Table of Contents

BBF RFC 105: Standardized Plant Syntax

October 28, 2014

1. Purpose

The difference in the biology of plants unveils a need for new standards. This Request for Comments (RFC) defines the common language for plant synthetic biology, extensible to all other Eukaryotes. Its syntax is based on the Golden Gate System1 cloning method which uses TypeIIS restriction enzymes; variations, such as GB2.02 and MoClo3, are already being used across the field. This RFC brings together their common features and sets a consensus across the plant field for construct assembly and part repositories.
Marchantia polymorpha, a primitive liverwort, is being established as a new model organism for plant Synthetic Biologyi4; this RFC uses M. polymorpha as well as Nicotiana benthamiana and Arabidopsis thaliana, two widely used model plant species, as exemplar chassis.

The fewer the rules defining a standard, the more widely that standard is adopted. The introduction of this new chassis with its novel tool kit is the occasion to allow more flexibility in the Registry5, opening it to wider usage and contributions.

[1] Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, et al. (2014) A golden gate modular cloning toolbox for plants. ACS Synth Biol: doi:10.1021/sb4001504.
[2] Sarrion-Perdigones A., Falconi E.E., Zandalinas S.I., Juarez P., Fernandez-del-Carmen A., Granell A., Orzaez D. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 2011;6:e21622.
[3] Weber E., Engler C., Gruetzner R., Werner S., Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PLoS One 2011;6:e16765.
[4] http://synbio.org.uk/marchantia/
[5] http://parts.igem.org/Main_Page

2. Relation to other BBF RFCs

The advantages of Type IIS restriction enzymes has been highlighted in RFC28, RFC53 and RFC61. RFC88 proposed to establish the MoCloGolden Gate standards in the *S. cerevisiae* chassis. RFC92 suggested the PartsRegistry to include parts compatible with both BioBricks and Golden Gate assembly standard. This RFC proposes to bring together the variations of the Golden Gate standard and establish a common language across the field.

3. Copyright Notice

4. A Common Type IIs Grammar for Eukaryotic transcriptional units

The Golden Gate system, based on Type IIS restriction enzymes, allows for one-shot assembly of multiple parts.

i. Introduction to the Golden Gate Standards

i. The Golden Gate cloning is a module-based cloning technique based on Type IIS restriction enzymes that enables parallel assembly of multiple parts in a one-pot, one-step reaction.
ii. Contrary to Type II restriction enzymes, Type IIS restriction enzymes cleave downstream of their recognition site. This enables a single enzyme to produce user-defined overhangs that we call fusion sites.
iii. The Golden Gate Modular Cloning Grammar sets standard flanking fusion sites for parts of same nature

[6] http://biobricks.org/

- e.g.: promoters, terminators. The grammar is set in such a way that the terminating fusion site of a part (e.g.: promoters) is identical to the starting fusion site of the next part (e.g.: coding sequence). This enables full transcriptional units to be constructed from a number of standard constitutive parts with a single enzyme in a single reaction.

iv. Golden Gate assembly bypasses time consuming steps of cloning such as custom primer design and order, PCR amplification and gel purification. Contrary to most assembly methods, it does not require overlapping flanking sequences but only four base pair fusion junctions, which can be made to be scarless if necessary.

v. The only requirement for implementing Golden Gate cloning is to purchase the required Type IIS restriction enzymes and to remove the recognition sites for those enzymes from the parts and the accepting backbone into which parts will be assembled. In this standard we define the use of three type IIS restriction enzymes: BsaI, BsmBI and Bpil. These recognise 6 bp sequences, therefore the recognition sites occur relatively sparsely throughout genomes:
BsaI GGTCTC$(N)_{1_5}$
BsmBI CGTCTC$(N)_{1_5}$
Bpil (BbsI) GAAGAC$(N)_{2_6}$

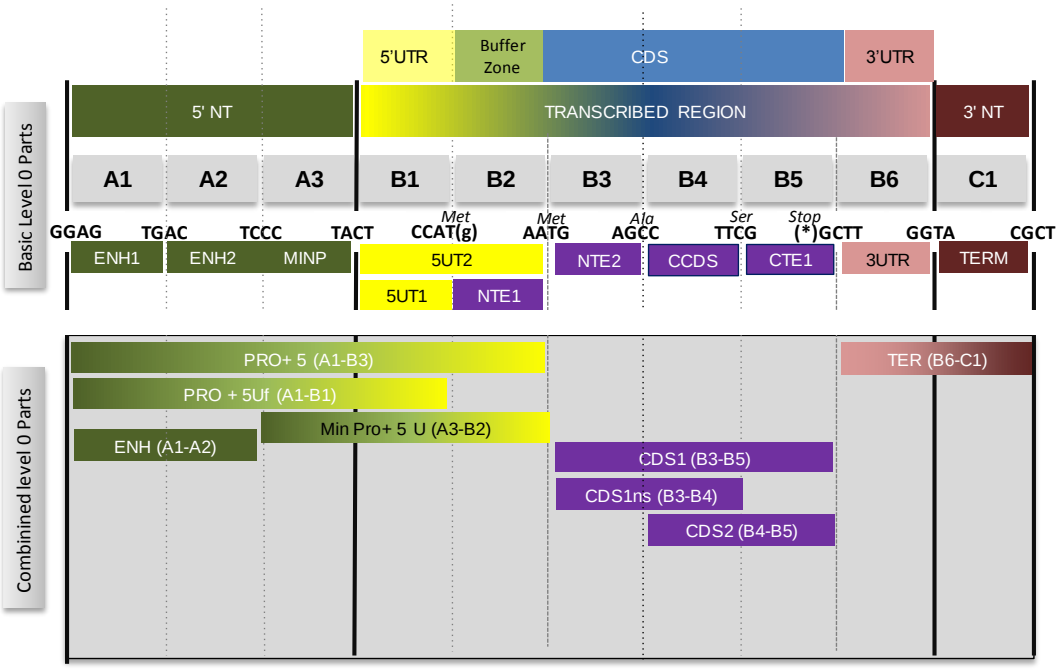
vi. Golden Gate assembly is optimal for constructing large combinatorial libraries. Parts of the same nature are flanked by the same two 4-bp overhang sequences making them fully interchangeable. This allows for large automated screens.

ii. Part Standards

i. A common grammar has to be established to enable sharing of parts throughout the community. This includes agreeing on the nature of parts and the fusion sequences allocated to them.

ii. Ideally the 4-bp overhangs include sequences that occur in all parts of same nature e.g.: the start codon atthe beginning of all coding parts. This allows minimal scars between parts, in contrast to the obligatory 6 bp scar sequence associated with BioBrick assembly.
iii. We refer to the removal on internal sites and the addition of the fusion sequences as domestication.
iv. Here we define the nature of parts and their fusion sequences.

Standard grammar



Basic Level 0 parts

BASIC STANDARD POSITION	*BASIC CATEGORY*	*FUNCTION*	*5' overhang*	*3' overhang*
A1	**ENH1**	Operator, cis regulatory region or transcriptional enhancer	GGAG	TGAC
A2	**ENH2**	Operator, cis regulatory region or transcriptional enhancer	TGAC	TCCC
A3	**MINP**	Minimal promoter, including transcription start	TCCC	TACT
B1	**5UT1**	5´-UTR region without obligate Met_i at the end of the fragment	TATC	CCAT
B2	**NTE1**	N-terminal coding region without obligate Met_i at the beginning of the fragment	CCAT	AATG
B1B2	***5UT2***	*Alternative 5´-UTR region with obligate Met_i at the end of the fragment*	*TACT*	*AATG*
B3	**NTE2**	Alternative N-terminal coding region with obligate Met_i at the beginning of the fragment	AATG	AGCC
B4	**CCDS**	Central Coding region without Met_i, without Stop codon.	AGCC	TTCG
B5	**CTE1**	C-terminal coding region with obligate Stop codon at the end of the fragment	TTCG	GCTT
B6	**3UTR**	3´- Untranscribed region	GCTT	GGTA
C1	**TERM**	Transcription terminator and polyA signal	GGTA	CGCT

Combined Level 0 parts

COMBINED STANDARD PARTS	*COMBINED CATEGORY*	*FUNCTIONAL ELEMENTS*	*PURPOSE*	*5′ overhang*	*3′ overhang*
A1-B2	PRO+5U	Full Promoter +transcription start + 5´UTR including Meti	Basic expression cassettes Translational Promoter fusions with NTE2 or CDS1	GGAG	AATG
A1-B1	PRO+5U(f)	Full Promoter including transcription start + 5´UTR without Meti	Basic expression cassettes Translational Promoter fusions with NTE1	GGAG	CCAT
A1-A3	PRO	Full promoter including transcription start, without 5´UTR	Transcriptional Promoter fusions with B1	GGAG	TACT
A3-B2	MinPRO+5UTR	Minimal promoter (TATA box) including transcription start, 5´UTR and Meti	Minimal promoter fusions with CDS1	TCCC	AATG
B1-B2	5UT1	Alternative 5´-UTR region with obligate Met_i at the end of the fragment	*Translational fusions with NTE2 and CDS1*	TACT	AATG
B3-B5	CDS1	Full coding region including Meti and Stop codon	Basic Expression cassetes	AATG	GCTT
B3-B4	CDS1ns	Coding region including Meti without stop codon	Protein Fusions with C-t domains	AATG	TTCG
B4-B5	CDS2	Coding region without Meti aith stop codon	Protein Fusions with signal peptides or other N-t domain	AGCC	GCTT
B6-C1	3´+ TERM	Combined regulatory regions including Stop codon, 3´UTR and Transcription Terminator	Terminator fusions for basic expression cassettes	TTCG	GCTT

iii. Common Vector for Level 0 Parts

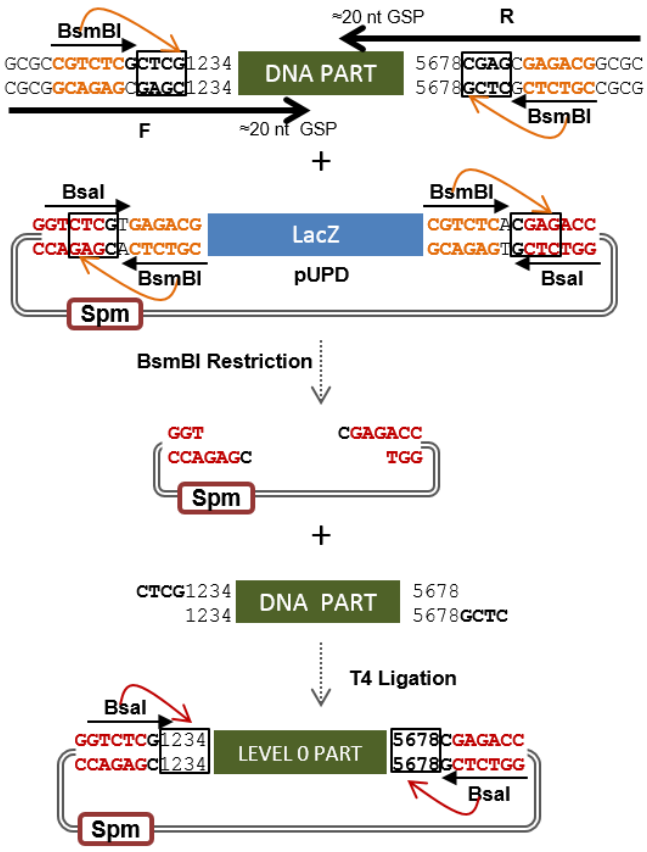
i. Each individual domesticated part is cloned into a universal level 0 plasmid- pUDP containing a spectinomycin resistance gene (SpecR).
ii. When cloned into this vector the part is flanked by two BsaI sites in opposing orientations: 5’ GGTCTCa_nnnn_PART_nnnn_tGAGACC 3’ with ‘nnnn’ being the part specific fusion sequence and ‘a’,’t’ nonspecific nucleotide.
iii. Parts are released from the Level 0 vector with a BsaI digest.

iv. Domestication Protocol of Level 0 Parts

i. Level 0 parts are domesticated into a pUDP (Universal Domesticator Plasmid) by amplifying the part with primers that include a *Bsm*BI recognition site followed by the four bp fusion sequence and approximately 20 gene specific nucleotides. 5'- AAAA**GGTCTCa**GGAG - promoter_region - AATG**tGAGACC**TTTT.
ii. Level 0 acceptor pUDP contain outward facing BsmBI sites into which the part will be received and can either contain inward facing BsaI sites for further use, or an overhang such that the BsaI site is generated upon ligation.
iii. The amplified part and acceptor vector are brought together with a BsmBI restriction ligation reaction and as a result the BsmBI sites are lost. The resulting

vector and part are releasable with the correct fusion sites by BsaI.

Domestication method of Level 0 Parts



5. Golden Gate Assembly Rules

In GoldenGate multipartite reactions, level 0 parts are assembled into higher order functional structures, primarily Transcriptional Units (TUs). Standard Level0 parts are released from pUDP vector and joined together in a multipartite fashion using a restriction/ligation reaction

catalyzed by BsaI and T4 ligase. Goldengate multipartite assemblies are governed by a single general assembly rule:

pUDP1(GGAG, i) +.... pUDPm(i, j)....+pUDPn(j, CGCT) + pDP = pEP(X)

where,

i. pUDP1(GGAG, i) .... pUDPm(i, j)....pUDPn(j, CGCT) is a ordered string of 1≤ n ≤10 standard level 0 parts,
ii. pUDP1(GGAG, i) is a level 0 part containing a 5' GGAG overhang site and any standard 3´overhang site (Figure 1),
iii. pUDPn(j, CGCT) is a level 0 part containing any standard 5´ overhang site (Figure 1) and a 3' CGCT overhang site,
iv. pUPDm(i, j) is a standard level 0 part whose overhang sites (i, j) are standard (Figure 1) and compatible with adyacent parts located upstream and downstream in the string, respectively.
v. pDP (destination plasmid) is a standard T-plasmid for plant transformation (see section 7) that generates GGAG and CGCT overhangs upon BsaI digestion and contains a selection marker that is not Spectinomycin.
vi. pEP(X) is a standard T-plasmid, derived from pDP, which harbors the composite DNA part X that results from the ordered fusion of the n standard parts described above, separated by (i,..j) 4 nt seams.

6. Multigenic constructs

Multigene assemblies are employed to pack multiple genetic instructions in a single T-plasmid. If appropriate T-plasmids (pDPs) are used, Golden Gate-assembled composite parts (usually TUs), can be easily combined to create multigene constructs. Here we describe two assembly methods that perform this operation: MoClo and GoldenBraid.

a. GoldenBraid assembly

The GoldenBraid (GB) system makes use of two types of T-plasmids as destination plasmids; α destination plasmids have BsaI sites facing outwards and BsmBI sites facing inwards, whereas Ω destination plasmids have inwards facing BsaI sites and outwards facing BsmBI sites.

In order to be reusable for GB multigene assemblies, Golden Gate-made TUs need to be assembled into one of two possible α destination plasmids (named pDGBα1 or pDGBα2). Similarly as described for the general Golden Gate assembly rule, GoldenBraid reusable TUs can be constructed in a BsaI/T4 ligase restriction/ligation reaction following any of the two operations described below:

1. pUDP1(GGAG, i) +.... pUDPm(i, j)....+pUDPn(j, CGCT) + pDGBα1 = pEPα1(X)

2. pUDP1(GGAG, i) +.... pUDPm(i, j)....+pUDPn(j, CGCT) + pDGBα2 = pEPα2(X),

where

i. pUDP1(GGAG, i), pUDPm(i, j), pUDPn(j, CGCT) are the same standard parts described for the Golden Gate general assembly rule.

ii. pDGBα1 is a KanR destination vector that, in addition to the general pDP vector features, contains BsmBI sites facing inwards and generates GGAG and GTCA overhangs upon BsmBI digestion.

iii. pDGBα2 is a KanR destination vector that, in addition to the general pDP vector features, contains BsmBI sites facing inwards and generates GTCA and CGCT overhangs upon BsmBI digestion.

iv. pEPα1(X) and pEPα2(X) are standard T-plasmids, derived from pDGBα1 and pDGBα2 respectively, which

harbor a composite DNA part (X) that results from the ordered fusion of the n standard parts described above, separated by 4 nt seams.

Next, GB multigene constructs can be assembled binarely following the iterative operations below:

1. pEPα1(Xi) + pEPα2(Xj) + pDGBΩ1 = pEPΩ1(Xi+Xj)
2. pEPα1(Xi) + pEPα2(Xj) + pDGBΩ2 = pEPΩ2(Xi+Xj)
3. pEPΩ1(Xi) + pEPΩ2(Xj) + pDGBα1 = pEPα1(Xi+Xj)
4. pEPΩ1(Xi) + pEPΩ2(Xj) + pDGBα2 = pEPα2(Xi+Xj),

where

i. (Xi), (Xj) and (Xi+Xj) are composite parts, comprising a single TU or a multigene structure, that have been created following any of the GoldenBraid assembly rules.

ii. pDGBα1, pDGBα2, pEPα1(Xi) and pEPα2(Xj) are the same elements described for the assembly of reusable GoldenBraid composite parts.

iii. pDGBΩ1 is a non-KanR pDP vector that contains BsmBI sites facing outwards and BsaI sites facing inwards and generates GGAG and GTCA overhangs upon BsaI digestion.

iv. pDGBΩ2 is a non-KanR pDP vector that contains BsmBI sites facing outwards and BsaI sites facing inwards and generates GTCA and CGCT overhangs upon BsaI digestion.

v. pEPΩ1(Xi) and pEPΩ2(Xi) are standard T-vectors, derived from pDGBΩ1 and pDGBΩ2 respectively, which harbor a composite DNA part (Xi) that results from the ordered fusion of n standard parts, separated by 4 nt seams.

vi. pEPα1(Xi+Xj), pEPα2(Xi+Xj), pEPΩ1(Xi+Xj) and pEPΩ2(Xi+Xj) are standard T-vectors, derived from

pDGBα1, pDGBα2, pDGBΩ1 and pDGBΩ2 respectively, which harbour the composite DNA part (Xi+Xj) resulting from the ordered fusion of (Xi) and (Xj) composite parts, where (Xi+Xj) follows the same assembly rules as (Xi) and (Xj).

vii. Operations 1 and 2 are performed using BsmBI and T4ligase restriction/ligation reaction.

viii. Operations 3 and 4 are performed using BsaI and T4ligase restriction/ligation reactions.

b. MoClo assembly

7. Unified features of binary vector for agrobacterium-mediated delivery

a. Introduction to Agrobacterium-mediated delivery

i. *Agrobacterium tumefaciens* is a soil phytopathogen that naturally infects plants by injecting its 'transferred (T)-DNA' into the host plant genome causing tumour-like crown gall disease. The T-DNA is delivered via the bacterial type IV secretion system. This ability has established *A. tumefaciens* as one of nature's own genetic engineers.[7]

ii. The molecular transfer of DNA is dependent on the expression of virulence genes resident on the tumour-inducing (Ti) plasmid. The T-DNA is defined and recognized by conserved, flanking T-DNA borders known as the Left (LB) and Right Border (RB). These polarized border sequences serve as the target for the Vir

[7] Gelvin, Stanton B. "Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool." *Microbiology and molecular biology reviews* 67.1 (2003): 16-37.

endonucleases which subsequently assist in integration of the T-DNA in the plant genome.

iii. For ease of engineering, scientists have separated the virulence genes and the T-DNA on separate plasmids to create a binary vector system: consisting of a vir-helper plasmid containing the virulence genes and T-plasmid containing the T-DNA flanked by the conserved borders. The T-plasmid commonly contains an additional origin of replication for E.coli for easy replication and assembly of the T-DNA region. T-regions as large as 200 kbp have been successfully mobilized[8]with increasing efficiency with smaller T-regions.

iv. Agro-mediated transformation is now an established method of plant transformation with variations depending on its usage: transient expression or stable integration.

v. We outline here the common features of the T-plasmid that are shared throughout the methods.

b. Standard Features for Agrobacterium-mediated delivery backbones

i. The T-plasmid comprises a conserved backbone with origins of replication for *E. coli* and *A. tumefaciens* and an antibiotic resistance gene cassette for selection in bacteria.

ii. It also contains a region to be replaced by the T-DNA variable sequence, flanked by the LB and RB sequences.

iii. Within the T-DNA, just next to LB and RB are outward facing BsaI sites to allow Type IIS restriction mediated insertion.

iv. To allow for selection in non-transient systems, a plant induced resistance cassette is included as part of the T-DNA. Insertion of T-DNA is directional with incorporation starting with the RB and progressing to the left. To ensure complete integration of T-DNA it is therefore advisable to

[8] Miranda, A., G. Janssen, L. Hodges, E. G. Peralta, and W. Ream. 1992. Agrobacterium tumefaciens transfers extremely long T-DNAs by a unidirectional mechanism. J. Bacteriol. 174:2288-2297.

place the antibiotic cassette adjacent to the LB. This increases the likelihood that resistant plants contain the complete T-DNA region.

v. The selection cassette usually consists of an antibiotic or herbicide resistance coding sequence driven by a strong constitutive or early-induced promoter but other selection cassettes such as fluorescent reporters are also used in some plant species[9].

[9] A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana.
Shimada TL, Shimada T, Hara-Nishimura I.
Plant J. 2010 Feb 1;61(3):519-28. doi: 10.1111/j.1365-313X.2009.04060.x. Epub 2009 Nov 25.
Erratum in: Plant J. 2010 Jul;63(2):352.

8. Applications

i. This section outlines examples where the PlantSyntax standard can offer significant benefit to the utilizer and community as a whole. We discuss applications to large-scale library assays and to the construction of versatile reusable expression cassettes.

a. Library construction

i. To engineer sophisticated circuits, a wide range of characterised elements are required. Advances in genome and transcriptome sequencing can provide large lists of such candidate regions. In order to yield useful parts, these *in silico* candidates need to be assayed *in vivo*. The PlantSyntax Standards and Type IIs mediated system enable a quick and reliable way to extract, test and characterize candidates. This is illustrated below by an outline of a promoter screen conducted in *M.polymorpha*.

ii. Candidate promoter regions in the *M. polymorpha* genome were identified by a phylogenetic comparison of predicted ORFs with the genome of *A. thaliana* and *P. patens*.

iii. Using primers described in 'domestication protocol of level 0 parts', DNA was extracted and amplified from genomic DNA with BsaI flanking sites enabling a subsequent one step insertion in acceptor plasmids.

iv. The same amplicon can be ligated in a range of different acceptor plasmids for direct *A. tumefaciens* mediated transformation enabling rapid optimisation of the assay by testing a range of reporters and expression systems.

v. The relatively small length of the oligonucleotides and need of only BsaI and ligase provide significant cost advantage

over other assembly techniques, such as Gibson[10], for conducting high-throughput assays.

b. Expression Cassettes

i. For part characterization and circuit optimization, a range of elements of the same nature must be incorporated and tested. Standardized expression cassettes allow users to plug, assay and compare their components systematically allowing synchronous evaluation of different parts in parallel.

9. *Marchantia polymorpha*, a model organism

i. Introduction to the new chassis

i. Liverworts are descendants of the earliest terrestrial plants. Their small and modular morphological characteristic is a reflection of their simple genomic structures. Recent developments in transformation techniques[11] can yield within 13 days a large number of stably transformed plants. This ease of transformation, combined with an increasing amount of genome characterisation place *M. polymorpha* as an ideal experimental chassis.[12] Contrary to historical model species, the choice of *M. polymorpha* has preceded its establishment, allowing the community to set standards in a coherent and united way paving the way for faster and more efficient sharing of information and parts.

[10] Gibson D.G., Young L., Chuang R.Y., Venter J.C., Hutchison C.A. 3rd., Smith H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 2009;6:343-345.
[11] AgarTrap: a simplified Agrobacterium-mediated transformation method for sporelings of the liverwort Marchantia polymorpha L.
Tsuboyama S1, Kodama
Y. Plant Cell Physiol. 2014 Jan;55(1):229-36. doi: 10.1093/pcp/pct168. Epub 2013 Nov 19.
[12] Marchantia.org

ii. Marchantia OpenParts Library and Codon optimization

i. OpenParts is the new library of *M. polymorpha* optimized parts. This repository is being populated and characterized with species specific promoters, enhancers and terminators. Using the optimized codon table of Marchantia Polymorpha, pre-existing parts would also be integrated.
ii. A codon usage table of *M. polymorpha* is being elaborated based on its genome and transcriptome sequence data. Codon optimization has proved to be critical for the expression of certain proteins in non-native chassis; and with decreasing synthesis costs, codon optimization is becoming an increasingly more feasible way of improving expression.

Codon Usage Table of *Marchantia polymorpha- to be changed xxxxxxxxxxx*

iii. Marchantia Agoro-mediated Transformation protocol and T-plasmid specification

i. Stable integration of foreign DNA in *M.polymorpha* can be achieved in 10 days and selected by 13.
ii. *M.polymorpha* spores are sterilized and plated. The purified T-plasmid is electroporated into *A. tumefaciens. Transformed A.tumefaciens* are selected on LB-agar plates containing the appropriate antibiotics at 30°C for 2-3 days. Isolated colonies are used to start 10 mL cultures in LB liquid medium containing the appropriate antibiotics and grown for 24h at 28°C. The cultures are centrifuged at 3000 g for 10 min and the pellet is re-suspended in approximately 2 mL (adjusted to OD600 1.5-2) of 1/2 GB media, 5% sucrose, 100µM acetosyringone and incubated for six hours with agitation. The now 5 day old spores are scraped from the plates in 25ml ½ GB, 5% sucrose and 100µM acetosyringone solution and

cocultivated with 1ml of the Agro solution for 1.5 days with agitation under full illumination at RT. The spores are collected with a 40µm cell strainer, washed with 150mL sterilized ddH20 and 100µg/mL cefotaximine and plated on ½ GB media with 10µg/mL hygromycin + 100µg/mL cefotaximine. *M. polymorpha* transformants can be screened within 3 days.[13]

iii. Plantlets can be repropagated on 1/2GB media with the appropriate antibiotics, for accelerated growth.
iv. To produce isogenic lines, cells from gemmae cups, asexual reproductive organ present on thalli from 4 weeks on[14], are collected and plated on selective media.
v. For selection of stable transformants in *M. polymorpha*, a hygromycin resistance cassette is included adjacent to the LB.
vi. The hygromycin resistance gene is followed by the 3' untranslated sequence and terminator from the nopaline synthase gene (nos) from *A. tumefaciens* and promoted by MPEF1α. MPEF1α is an endogenous *M. polymorpha* promoter active from early in development, and its use can lead to a high transformation efficiency.[15]
vii. Wild type spores are not vulnerable to standard concentrations of kanamycin so this should be avoided as a selection agent.[16]

[13] Protocol marchantia
[14] Effect of Light Intensity and Sucrose on the Production of Gemma Cups and Gemmae in Marchantia nepalensis R. N. Chopra and Sneh Sood
The Bryologist, Vol. 73, No. 3 (Autumn, 1970), pp. 592-596
[15] Comparison of the MpEF1α and CaMV35 promoters for application in Marchantia polymorpha overexpression studies. Althoff F, Kopischke S, Zobell O, Ide K, Ishizaki K, Kohchi T, Zachgo S.
Transgenic Res. 2014 Apr;23(2):235-44. doi: 10.1007/s11248-013-9746-z. Epub 2013 Sep 15.
[16] Unpublished kana info

10. *Nicotiana benthamiana* and Arabidopsis thaliana: model dicotyledenous plants

i. Description of the chassis

i. *N. benthamiana* is a widely used experimental plant from the solanaceous group of flowering plants that includes tomatoes, potatoes and capsicums. It is widely used in plant pathology due to the large number of plant pathogens (viruses, bacteria, fungi, oomycetes etc) that can successfully infect it. Of importance to synthetic biologists, *N. benthamiana* it is easily genetically transformed and regenerated and amenable to facile methods for virus-induced gene silencing and transient protein expression including the production of therapeutic compounds and pharmaceuticals.

ii. *Arabidopsis thaliana* is a member of the mustard (*Brassicaceae*) family that includes species such as oilseed rape, cabbage and turnip and is very widely established as a model organism in plant biology. It became popular as model organism due to its small stature, short lifecycle, amenability to genetic manipulation and transformation and its small diploid genome, for which a sequence was published in 2000 (The Arabidopsis Initiative, Nature 2000) A large number of mutant lines and genomic resources are available at the TAIR database (The Arabidopsis Information Resource)[17].

ii. GoldenGate parts and codon-optimisation for model dicotyledenous plants

i. A Golden Gate resource for plant biologists, comprising a set of cloning vectors for the domestication of new parts and 96 standardized parts to enable facile construction of multigene constructs

[17] http://www.arabidopsis.org

for plant transformation was published in 2014 (Engler et al 2014). Parts include promoters, untranslated sequences, reporters, antigenic tags, localization signals, selectable markers, and terminators. The comparative performance of regulatory parts was performed in *N. benthamiana.* These tools are available from the authors and the AddGene repository[18].

ii. Although only a handful of nuclear genes of higher plants have been sequenced, a clear distinction between the codon usage of dicotyledenous and monocotoledenous plants is evident. Species specific codon optimization has proved to be beneficial expression of certain proteins in non-native chassis; and it is recommended to codon-optimise if data is available. However, with limited transcriptome data for a large number of plant species, the most widely used codon usage table for dicotyledenous plants is that of *A. thaliana* based on 33661 coding sequences:

TTT 22.7(331889)	TCT 25.0(366199)	TAT 15.2(223175)	TGT 10.9(160132)
TTC 20.3(297191)	TCC 10.9(159324)	TAC 13.5(198197)	TGC 7.2(105486)
TTA 13.2(193377)	TCA 18.3(267571)	TAA 0.9(12660)	TGA 1.0(14350)
TTG 21.3(311721)	TCG 9.0(132078)	TAG 0.5(6985)	TGG 12.6(184264)
CTT 24.2(354235)	CCT 18.3(267956)	CAT 14.1(206737)	CGT 8.8(128993)
CTC 15.7(230429)	CCC 5.2(76715)	CAC 8.6(126093)	CGC 3.7(54377)
CTA 10.1(148530)	CCA 16.1(235874)	CAA 19.7(288312)	CGA 6.3(92902)
CTG 10.0(146833)	CCG 8.2(120792)	CAG 15.0(219990)	CGG 4.8(70654)
ATT 22.1(323223)	ACT 17.6(257393)	AAT 23.1(338400)	AGT 14.3(210058)
ATC 18.3(267935)	ACC 10.1(148145)	AAC 20.7(303888)	AGC 11.2(163949)
ATA 13.1(191439)	ACA 15.9(233235)	AAA 31.3(458631)	AGA 19.0(278669)
ATG 24.4(357938)	ACG 7.6(111358)	AAG 32.5(476138)	AGG 10.9(160105)
GTT 27.2(398603)	GCT 27.5(402327)	GAT 37.2(544870)	GGT 21.7(317299)
GTC 12.5(183353)	GCC 10.0(145775)	GAC 17.1(250119)	GGC 8.9(130312)
GTA 10.2(149877)	GCA 17.4(255578)	GAA 35.0(512425)	GGA 23.5(344625)
GTG 17.3(252815)	GCG 8.6(126294)	GAG 32.1(470306)	GGG 10.1(148212)

[18] https://www.addgene.org/

iii. Agrobacterium-mediated Transformation protocol

i. Transient expression of foreign DNA can be detected within 28 hours of transfection of young *N. benthamina* leaves infiltrated with *A. tumefaciens. A. tumefaciens* are selected on LB-agar plates containing the appropriate antibiotics at 28°C for 2-3 days. Isolated colonies are used to start 5 mL pre-cultures in LB liquid medium containing the appropriate antibiotics and grown for 24h at 28°C. A larger culture is made by adding 500µL of pre-culture to 50 mL LB liquid medium with the appropriate antibiotics. The flask is incubated for 16 hours at 28°C with agitation. The bacterial cultures are centrifuged at 3200 g for 10 min to pellet the cells. The supernatant is discarded and each pellet is re-suspended in 5 mL of MMA (10mM MgCL2, 10mM MES/KOH pH 5.6; 150µM acetosyringone). The cells are incubated in MMA for two hours in the dark at room temperature. The optical density of the culture at 600 nm is measured with a spectrophotometer and adjusted to 0.8 by diluting with extra MMA. For the co-infiltration of multiple *A. tumefaciens* cultures, equal volumes of the 0.8 OD_{600nm} *A. tumefaciens* cultures in MMA are mixed together.Prior to infiltration, holes are pierced in the underside of leaves of 3 weeks old *Nicotiana benthamiana* plants using a sterile pipette tip. One millilitre needle-less syringes are then used to infiltrate the entire leaves with the *A. tumefaciens* cultures. Tissues are collected approximately 6 days after agro-infiltration (depending on the genes expressed) and can be directly used for analysis, freeze-dried or frozen in liquid nitrogen prior to storage at -80°C.

ii. For stable integration of the T-DNA, pieces of leave are sterilised and co-cultivated with *A. tumefaciens* cultures in the dark for 48 hours. The pieces of leaf are then transferred to selection media where they form

callus at the cut edges. Resistant callus are removed to media containing hormones to induce shoot elongation and, subsequently, resistant shoots are moved to media to allow the production of roots. Plantlets that form roots are transplanted to soil.

iii. Stable transformation of *A. thaliana* is achieved by a floral dip method. Plants are grown in optimal conditions until they are flowering. Optimal plants should have multiple immature flower clusters should not have started to set seed. *A. tumefaciens* cultures are grown in liquid culture to OD_{600} = 0.8 and resuspended in a 5% sucrose solution. 200ml are required for each 9cm pot to be dipped. Dip above-ground parts of plant in Agrobacterium solution for 2 to 3 seconds and then leave the pot on its side in a large plastic beaker or under a dome to maintain humidity for 24 hours. Water and grow plants normally until the seeds mature. Dried seed pods are harvested and the seed released from the pods. Seeds should be sterilized and germinated on 0.5X MS/0.8% containing the appropriate selection cassette or, alternatively, visualized under UV light for the presence of the seed-specific fluorescence conferred by a reporter cassette.

11. Additional Resources and Information

i. The OpenPlant[19] project has assembled on a single database the complete library of published research done on *Marchantia polymorpha*. This offers a connected platform of knowledge and a forum for scientific exchange. There are two strains of *polymorpha* under study. The Tokyo**strain's genome was published this year.[20] and is being compared to

[19] http://www.openplant.org/

[20] Tokyo genome

the CAM strain which is being sequenced and mined. There are currently 11 labs working on Marchantia system.[21]

12. Authors' Contact Information

13. References

i. B. A. Collins. Gibthon Construct Designer, October 2010. Available from: http://www.gibthon.org.

ii. http://synbio.org.uk/marchantia/

iii. The Arabidopsis Information Resource[22] (TAIR)

iv. The SolGenomics Network SolGenome[23]

[21] http://synbio.org.uk/marchantia/labs.html
[22] http://www.arabidopsis.org
[23]http://solgenomics.net